

Data Science in Action #8

“The Sales Managers’ Problem”

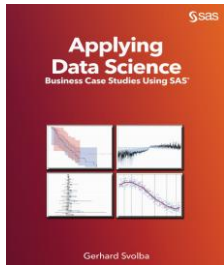
Using Monte Carlo Simulations to Understand the Outcome Distribution



Gerhard Svolba
Data Scientist, SAS Austria

Analytics and Data Science is there to help you!

- Get a clearer, more objective picture of your data and your analysis subjects
- Get explicit results instead of searching the needle in the haystack
- Make your data talk to you!
- Receive findings automatically instead of manually
- Do it again! – treat models as an asset and repeat your analysis

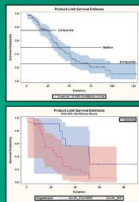


Data Science Applications and Case Studies

Data Science in Action: #1

Performing Headcount
Survival Analysis for
Employee Retention

*Can assumptions about the average
length of time intervals be made, even if
most of the endpoints have not yet been
observed?*



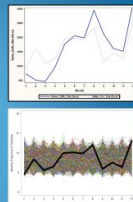
Survival analysis methods: Kaplan-Meier estimates
Cox Proportional Hazards regression
Survival Data Mining



Data Science in Action: #5

Checking the Alignment with
Predefined Pattern

*Which customers show a behavior that
is far from what you expected?*



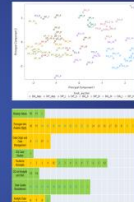
Chi2 independency test
Benford's law
Time Series Similarity



Data Science in Action: #7

Topic Search Documents and
Clustering

*Can I automatically find clusters of
documents with similar content?*



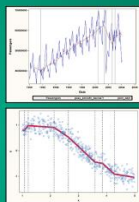
Text Mining
Text Parsing (Synonyme, Stemming, Stop-Listen)
Term by Document Weights



Data Science in Action: #2

Detecting Structural
Changes and Outliers in
Longitudinal Data

*Can events and changes in the
course over time be
automatically detected?*



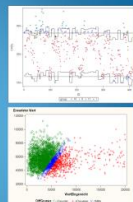
Smoothing Of Longitudinal Data
Multivariate Adaptive Regression Splines
Automatic Breakpoint Detection
Automatic Detection of Outliers with ARIMA Models



Data Science in Action: #6

Proving a reference value
that considers all available
co-information

*Can analytics help me to reduce the
"Yes, but ..." sentences in my business
discussions?*



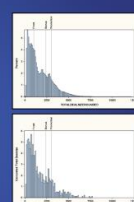
Linear Regression
Decision Trees
Time Series Analysis



Data Science in Action: #8

Using Monte Carlo Simulations to
Understand the Outcome
Distribution

*When the sales manager looks at the
project pipeline, does the sum of weighted
averages give him or her a full picture?*



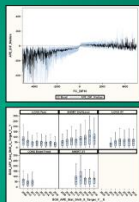
Monte Carlo Simulations
Mathematical Programming



Data Science in Action: #3

Explaining Forecast Errors
and Deviations

*Do the demand planners really improve
forecast accuracy with their manual
overwrites?*



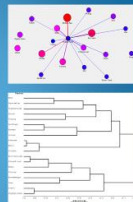
Linear Regression
Quantile Regression
Descriptive Statistics



Data Science in Action: #4

Listening to Your Data – Discover
Relationships with Unsupervised
Analysis Methods

*Can your data tell you stories about
your analysis subjects, even if you don't
ask explicitly?*



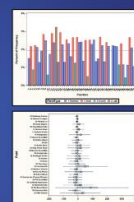
Unsupervised machine learning methods:
association analysis
variable clustering



Data Science in Action: #9

Studying Complex Systems –
Simulating the Monopoly Board
Game

*How can you simulate complex
environments to get insight in the most
frequent processes?*



Monte Carlo Simulations



Get access to more content:



SAS DACH @Youtube: <https://www.youtube.com/user/SASsoftwareGermany>

Blogs on LinkedIn: <https://www.linkedin.com/in/gerhardsvolba/>

Twitter: <https://twitter.com/gsvolba>

Content on Github: <https://github.com/gerhard1050>

Books @SAS-Press: <https://support.sas.com/svolba>

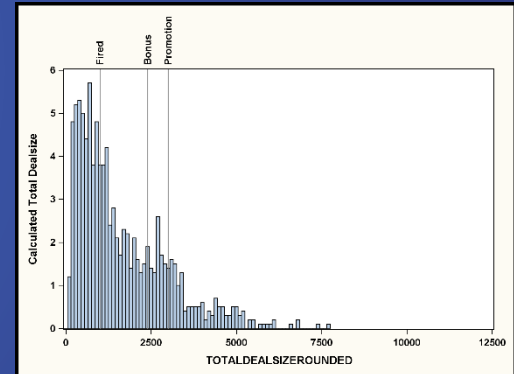
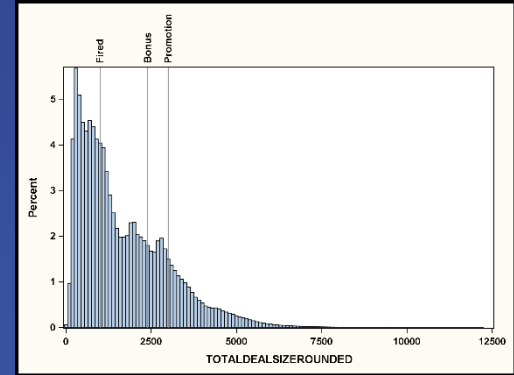


Data Science in Action: #8

Using Monte Carlo Simulations to Understand the Outcome Distribution

When the sales manager looks at the project pipeline, does the sum of weighted averages give him or her a full picture?

Monte Carlo Simulations
Mathematical Programming



Will the Sales Manager keep his Job?

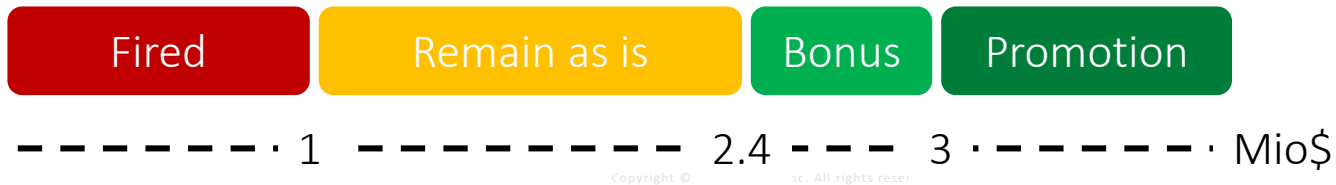
For each sales opportunity the sales manager has the project value and the probability of closing the deal.

ProjectID	DealSize (1000 \$)	Probability
1	1500	10%
2	10	65%
3	500	20%
4	50	50%
5	100	40%
6	30	90%
7	10	60%
8	150	20%
9	200	25%
10	180	10%
11	900	10%
12	750	20%
13	600	10%
14	320	20%
15	100	40%
16	50	80%
17	2000	5%
18	400	20%
19	2500	10%
20	1700	15%
21	100	80%

After his weekly phone call with his regional manager:

- If he manages to exceed 2,4 Mio in total, he will receive a special bonus
- If he manages to exceed 3 Mio in total, he will be promoted to the position of a regional director.
- If he achieves less than 1 Mio, he will most likely get fired.

Can the sales manager feel save?
Will he get a bonus or promotion?



Two ways to calculate that: A stupid and a simple one

ProjectID	DealSize (1000 \$)	Probability
1	1500	10%
2	10	65%
3	500	20%
4	50	50%
5	100	40%
6	30	90%
7	10	60%
8	150	20%
9	200	25%
10	180	10%
11	900	10%
12	750	20%
13	600	10%
14	320	20%
15	100	40%
16	50	80%
17	2000	5%
18	400	20%
19	2500	10%
20	1700	15%
21	100	80%

„Stupid Calculation": Deal Sum = € 12,15 Mio * AvgProb 33% = 4,05 Mio

Weighted mean (expected value): \$ 1,661,500

Fired

Remain as is

Bonus

Promotion

----- 1 ----- 2.4 ----- 3 ----- Mio\$

Want to have a clearer picture?

Some considerations for 2 projects

- Consider the case where the sales managers' project pipeline only contains the projects 3 and 5 from the list
- There are 4 potential scenarios that could happen:

Scenario	Value	Scenario Probability
Project 3 and 5	600.000	$0.2 * 0.4 = 0.08$
Project 3 only	500.000	$0.2 * 0.6 = 0.12$
Project 5 only	100.000	$0.8 * 0.4 = 0.32$
None of the projects	0	$0.8 * 0.6 = 0.48$

ProjectID	DealSize (1000 \$)	Probability
1	1500	10%
2	10	65%
3	500	20%
4	50	50%
5	100	40%
6	30	90%
7	10	60%
8	150	20%
9	200	25%
10	180	10%
11	900	10%
12	750	20%
13	600	10%
14	320	20%
15	100	40%
16	50	80%
17	2000	5%
18	400	20%
19	2500	10%
20	1700	15%
21	100	80%

Can we get this output for all 21 projects?

- No project is sold: 000000000000000000000000 0 \$
- Only project 1: 100000000000000000000000 1,500 \$
- ...
- Project 2,4,7: 020400700000000000000000 70 \$
- ...
- ...
- All projects are sold: 111111111111111111111111 12,15 \$
- In total 2,097,152 combinations for 21 projects!

ProjectID	DealSize (1000 \$)	Probability
1	1500	10%
2	10	65%
3	500	20%
4	50	50%
5	100	40%
6	30	90%
7	10	60%
8	150	20%
9	200	25%
10	180	10%
11	900	10%
12	750	20%
13	600	10%
14	320	20%
15	100	40%
16	50	80%
17	2000	5%
18	400	20%
19	2500	10%
20	1700	15%
21	100	80%

Can you perform such calculations automatically?

- YES! The SAS Viya platform allows you to perform such calculations within seconds.
- The code is very short:

```
proc iml;
  use projects; read all; close;

  N = nrow(prob);          /* number of projects */
  ScenarioID = t(1:2**N);
  format = "binary" + strip(char(N));
  bin = putn(ScenarioID,format);

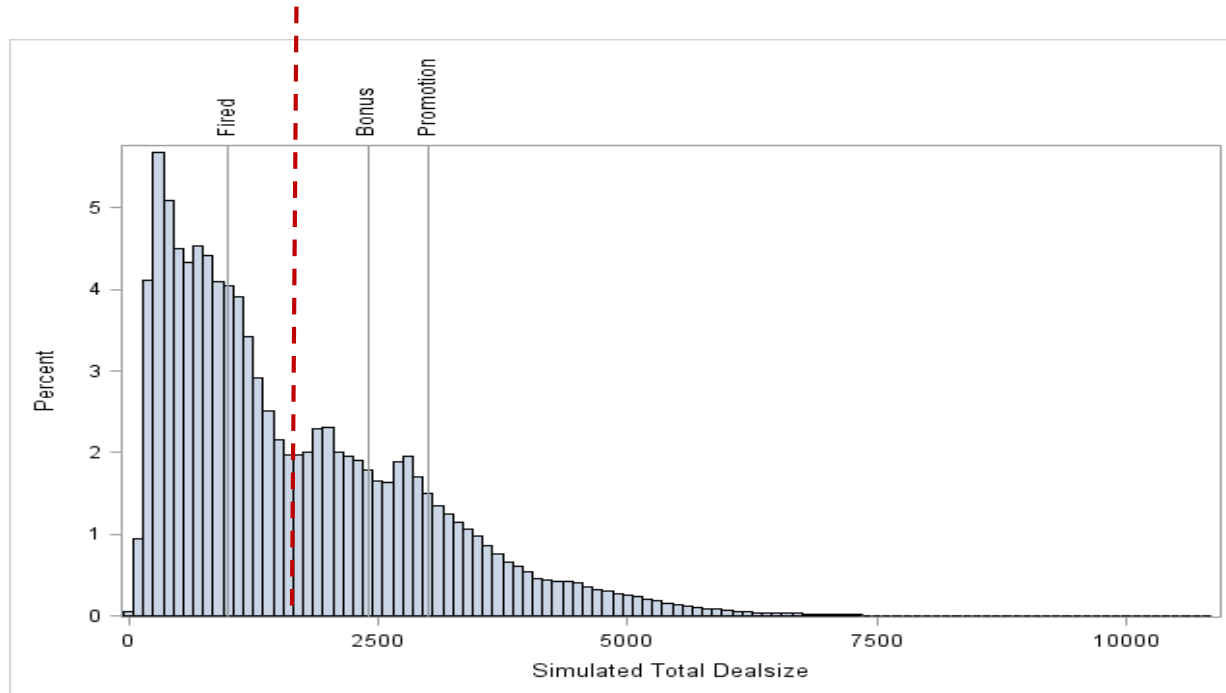
  bt=j(2**N,N,-1);
  do i = 1 to N;
    bt[,i]=num(substr(bin,N-i+1,1));
  end;
  prob_m = abs(1-bt-t(prob));

  ScenarioSum = bt * value;
  ScenarioProb = prob_m[,#];
  TotalDealsizeRounded = round(ScenarioSum,100);

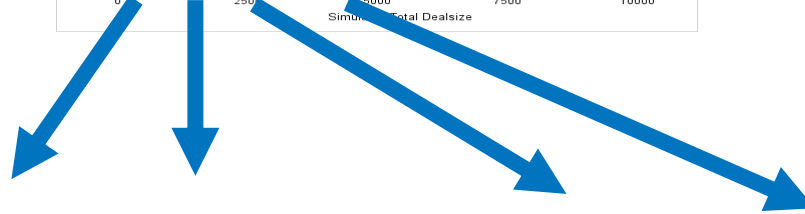
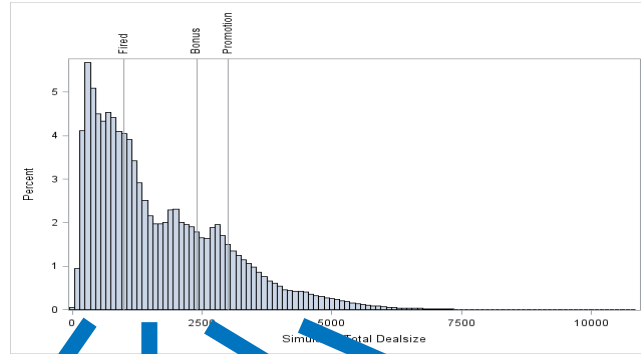
  create FullCalc_Outcomes_IML var {ScenarioID ScenarioSum TotalDealsizeRounded ScenarioProb};
  append;
  close FullCalc_Outcomes_IML;
quit;
```

Will the Sales Manager keep his Job?

Weighted Average:
\$ 1.661.500



Probabilities per Scenario



Full Calculation

40,36 %

33,94%

10,47 %

15,23 %

Fired

Remain as is

Bonus

Promotion

----- 1 ----- 2.4 ----- 3 ----- Mio\$

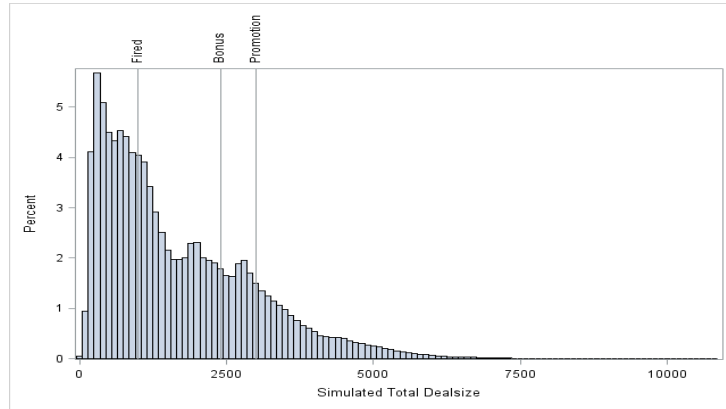
A different calculation approach: Monte Carlo Simulations

- Do we really need to perform all these many calculations or is there an easier way?
- Use Monte Carlo Simulations!
- How does that work?
- For each project toss a biased coin based on its probability
 - If YES record the \$ „project value“
 - If NO record \$ 0
- Do this for all projects and return the total project value as output.
- Repeat this procedure 1000s of times

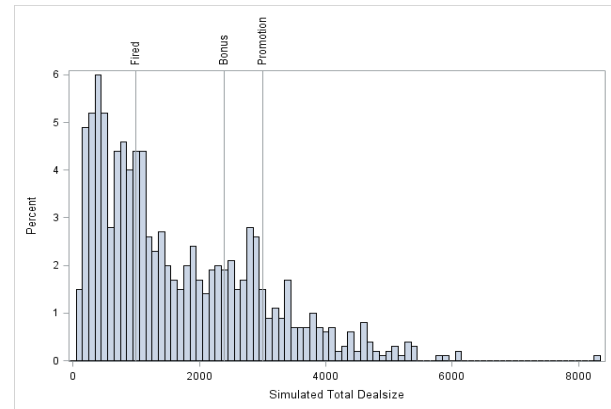
ProjectID	DealSize (1000 \$)	Probability
1	1500	10%
2	10	65%
3	500	20%
4	50	50%
5	100	40%
6	30	90%
7	10	60%
8	150	20%
9	200	25%
10	180	10%
11	900	10%
12	750	20%
13	600	10%
14	320	20%
15	100	40%
16	50	80%
17	2000	5%
18	400	20%
19	2500	10%
20	1700	15%
21	100	80%

Results from the Monte Carlo Simulations

1,000,000 Iterations



1,000 Iterations



Monte Carlo Simulations get very close to the true result!

1,000 Iterations	40,70 %	32,10%	12,20 %	15,00 %
1,000,000 Iterations	40,38 %	33,94%	10,44 %	15,24 %
Full Calculation	40,36 %	33,94%	10,47 %	15,23 %



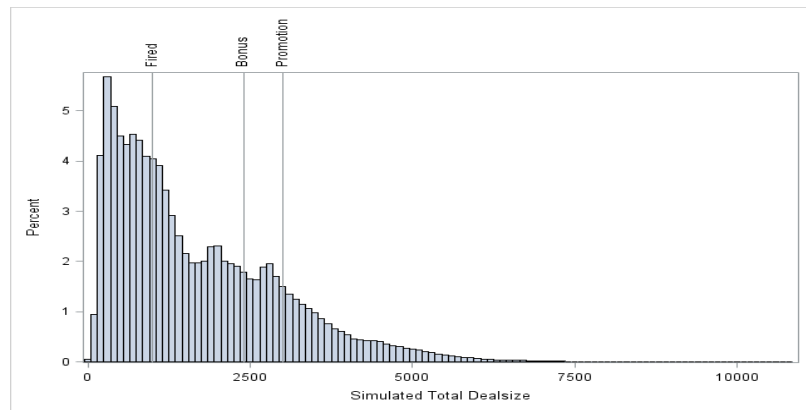
The code for complex analyses can be short and simple

```
data Scenario_Raw(drop=value prob);  
  call streaminit(19596);  
  set projects;  
  do ScenarioID = 1 to 1000000;  
    if rand('Uniform') < prob then OutcomeValue=value;  
    else OutcomeValue = 0;  
    output;  
  end;  
run;  
proc means data=Scenario_Raw noprint nway;  
  class ScenarioID;  
  var OutcomeValue;  
  output out= Scenario_Outcomes_Datastep(drop=_type_ _freq_)  
sum(OutcomeValue)=ScenarioSum;  
run;
```

Advantages of Monte Carlo Simulations

- The weighted average only provides you an incomplete picture.
- You want to study the distribution of the possible outcomes.
- Monte Carlo simulations are a powerful tool to perform such calculations.

Weighted Average:
\$ 1,661,500



Get access to more content:



SAS DACH @Youtube: <https://www.youtube.com/user/SASsoftwareGermany>

Blogs on LinkedIn: <https://www.linkedin.com/in/gerhardsvolba/>

Twitter: <https://twitter.com/gsvolba>

Content on Github: <https://github.com/gerhard1050>

Books @SAS-Press: <https://support.sas.com/svolba>

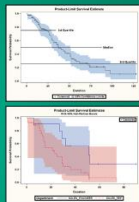


Data Science Applications and Case Studies

Data Science in Action: #1

Performing Headcount Survival Analysis for Employee Retention

*Can assumptions about the average
length of time intervals be made, even if
most of the endpoints have not yet been
observed?*



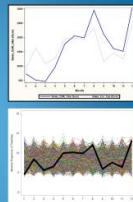
Survival analysis methods: Kaplan-Meier estimates
Cox Proportional Hazards regression
Survival Data Mining



Data Science in Action: #5

Checking the Alignment with Predefined Pattern

*Which customers show a behavior that
is far from what you expected?*



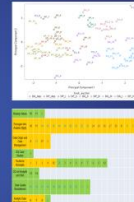
Chi2 independency test
Benford's law
Time Series Similarity



Data Science in Action: #7

Topic Search Documents and Clustering

*Can I automatically find clusters of
documents with similar content?*



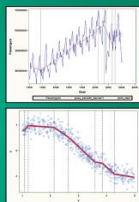
Text Mining
Text Parsing (Synonyme, Stemming, Stop-Listen)
Term by Document Weights



Data Science in Action: #2

Detecting Structural Changes and Outliers in Longitudinal Data

*Can events and changes in the
course over time be
automatically detected?*



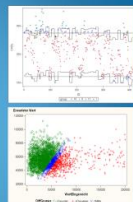
Smoothing Of Longitudinal Data
Multivariate Adaptive Regression Splines
Automatic Breakpoint Detection
Automatic Detection of Outliers with ARIMA Models



Data Science in Action: #6

Proving a reference value that considers all available co-information

*Can analytics help me to reduce the
"Yes, but ..." sentences in my business
discussions?*



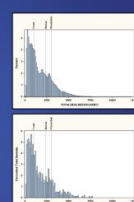
Linear Regression
Decision Trees
Time Series Analysis



Data Science in Action: #8

Using Monte Carlo Simulations to Understand the Outcome Distribution

*When the sales manager looks at the
project pipeline, does the sum of weighted
averages give him or her a full picture?*



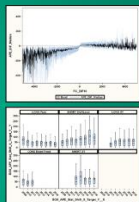
Monte Carlo Simulations
Mathematical Programming



Data Science in Action: #3

Explaining Forecast Errors and Deviations

*Do the demand planners really improve
forecast accuracy with their manual
overwrites?*



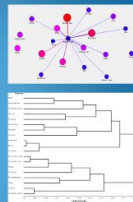
Linear Regression
Quantile Regression
Descriptive Statistics



Data Science in Action: #4

Listening to Your Data – Discover Relationships with Unsupervised Analysis Methods

*Can your data tell you stories about
your analysis subjects, even if you don't
ask explicitly?*



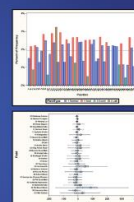
Unsupervised machine learning methods:
association analysis
variable clustering



Data Science in Action: #9

Studying Complex Systems – Simulating the Monopoly Board Game

*How can you simulate complex
environments to get insight in the most
frequent processes?*



Monte Carlo Simulations

